

AUTOMATIC SEGMENTATION OF CALCIFICATION AREAS IN DIGITAL BREAST IMAGES

JALAWI ALSHODUKHI & KHALID TWARISH ALHAMAZANI

University of Ha'il, College of Computer Science and Engineering, Department of Computer Science,
Kingdom of Saudi Arabia

ABSTRACT

In this study, the authors hope to demonstrate that when mammography is combined with intelligent segmentation techniques, it can become more effective in diagnosing breast abnormalities and aiding in the early detection of breast cancer. In conjunction with intelligent segmentation techniques, mammography can be made more effective in diagnosing breast abnormalities and aiding in the early diagnosis of breast cancer, hence increasing its overall effectiveness. The methodology, which includes some concepts of digital imaging and Machine Learning techniques, will be described in the following section after a review of the literature on breast cancer (categories, prevention involving the environment and lifestyle, diagnosis and tracking of the disease) has been completed (Neural Networks and Random Forests). It was possible to achieve these results by working with an image collection that previously had questionable regions (per the given technique). Fiji software extracted problematic candidate regions from mammography images, which were subsequently subjected to further examination. To categorize the results of the picture segmentation, they were sorted into three groups, which were as follows: Random Forest and neural networks both generated promising results in the segmentation of suspicious parts that were emphasized in the highlight of the image, and this was true for both algorithms. Detection of contours of the regions was carried out, indicating that cuts of these segmented sections may be created. Later on, automatic categorization of the targets can be carried out using a learning algorithm, as illustrated in the experiment.

KEYWORDS: SVM, Machine learning, artificial intelligence, Breast cancer, Image processing, & Diagnosis.

Received: Feb 05, 2022; **Accepted:** Feb 25, 2022; **Published:** Mar 08, 2022; **Paper Id.:** IJCNWMCJUN20223

1.0 INTRODUCTION

In Arab countries, the large number of women dying from breast cancer makes the disease a significant public health problem, both because of the commotion of morbidity and mortality, as well as the high personal and social cost related to the disease and its treatment, and even more because of the harmful physical and psychological squeal in patients [1]. It is a tumour that originates from breast cells that grow disorderly. 60% to 70% of this type of cancer appears in the form of an “irregular or spheroid” nodule, with speculated, indistinct or micro lobulated margins, without calcifications. Close to 20% of cases are nodules with calcifications. Calcifications without an associated nodule constitute just under 20% of all cases” [2].

Breast cancer is quite frightening because of its high frequency and the physical (either by total mastectomy or segmental resection – the scar left behind is essential) and psychological (low self-esteem, reduced sexuality, stigma) that generally harm patients. It is one of the diseases that most cause mortifications and disorders in a wide dimension: it affects the patient, close family members and caregivers [3].

The most significant risk of this disease is the late diagnosis of the tumour, which can be avoided with

digital technologies advancing at an ever-increasing pace to support the early diagnosis of the disease. Mammography, for example, is a test used to diagnose breast abnormalities; it is used in programs for screening women; it reduces the mortality rate due to breast cancer by more than 40% [4], obviously if the diagnosis is still in the pre-clinical stage, that is, in the early stage of the evolution of malignant tumours.

Because breast cancer is the second with the highest incidence globally and its prevention and mainly treatment are costly for the country's public health, Thresholding is a more uncomplicated implementation technique for image processing and computational speed. They are applied very frequently in the image segmentation scenario. It is also called binarization because the technique is based "on partitioning the histogram of the image to convert all pixels [picture element] whose grey tone is greater than or equal to a certain threshold value T into white and the others in blacks".[5]

That said, the question arises: among the many real possibilities of breast cancer prevention, what would be the relevance of the threshold in this effort? The search for answers to this question makes it possible to achieve the objective of this research, which is to discuss the advantages and limits of the threshold technique applied to mammography in the search for areas suspected of malignant neoplastic, facilitating the interpretation of the findings by the physician and, therefore, way, helping the early diagnosis of breast cancer.

1.1 Literature Review

Although there is no defined etiology, breast cancer is studied as multifactorial so that certain interacting risk factors determine the disease. Farhadihosseinabadi, et al.. (2020) [6] discuss the two categories of risk factors:

- **Modifiable** – subject to modifications, interventions, controls, such as "obesity, high-carbohydrate eating habits, exaggerated consumption of red meat and fats, high intake of alcoholic beverages, the performance of combined hormone replacement therapy for more than five years and excessive radiation exposure" [6]. It is read in Vieira that "overweight is considered a risk factor for the development of the disease and this can be explained by the high estrogen levels resulting from peripheral conversion in adipose tissue".
- **Non-modifiable** – they do not change. They are inherent to the patient, such as gender (higher incidence in females), age (predominance in those over 50 years of age), race and ethnicity (more frequent among non-Hispanic whites and blacks). Furthermore, [...] reproductive and hormonal status of women, with early menarche and late menopause one of the most associated, family history, association when first-degree relatives (mother, sister, daughter, father, brother and son) developed breast cancer. Mutation in the BRCA1 and BRCA2 genes [7], families carrying these mutations have strong indications for the disease and also the presence of previous breast pathology, women with previous breast cancer have a 1.5x increased risk of developing breast cancer again.

How to best prevent breast cancer? This is the question that should be part of the questioning of women, mainly because the answers would allow primary prevention of the disease, that is, before the beginning of the pathological process, even avoiding exposure to many modifiable risk factors, such as the environment and lifestyle, for example, thus preserving health and reducing mortality as a result of this disease.

Some habits should be part of every woman's life, namely: practising physical activities, breastfeeding an infant, regularly eating with fruits, vegetables, fish and nuts, olive oil and, on the other hand, avoiding the intake of fat and red meat (which is still not very clear), processed foods, foods with a high glycemic index, reducing the use of sugar [8].

Therefore, they are simple practices to be cultivated, especially when it is confirmed that, in Arab countries, the risk of having breast cancer is 8% throughout life, which means that one woman in twelve is an expressive risk female. Because of this, the American Society of Mastology, the American Society of Oncological Surgery, and the American Society of Radiology recommend screening the disease in women from the age of forty undergoing mammography annually [4] to detect any breast abnormalities.

Therefore, the early diagnosis of the pre-clinical stage, even before the presentation of any symptoms, considerably increases the chances of cure. There is secondary prevention, whose main objective is universal and early screening and the performance of mammography, applied "to large populations, in screening programs, significantly reduces mortality rates from breast cancer (reduction above 40%) " [6]

Secondary prevention aims to change the course of the disease since its biological onset has already occurred through interventions that allow its early detection and timely treatment. For this, there must be clear evidence that the disease in question can be identified at an early stage when it is not yet clinically apparent, allowing a practical therapeutic approach, altering its course or minimizing the risks associated with clinical therapy. Furthermore, the resulting drop in morbidity or mortality must be achieved without the adopted strategy's significant burden of adverse effects. Early detection of a disease is possible through education for early diagnosis in symptomatic people or screening (screening) in asymptomatic populations [9].

On the other hand, tertiary prevention occurs in a clinical and symptomatic phase of the disease in the face of findings such as a nodule, oedema – a phase in which mammography is no longer a screening to be diagnostic. "It is important to point out that mammography presents false positives and false negatives, an inherent flaw in the method, but it is the best screening method available at the moment"[10].

There are two groups of mammography exams: screening (in asymptomatic patients, primary prevention, which should be performed annually after age forty in the postmenstrual period) and diagnostic (in symptomatic patients suggestive of breast cancer or even those who need to be supplemented with another exam) [9].

Therefore, early diagnosis and treatment of the disease are essential so that the consequences are less harmful, avoiding or reducing mastectomies and the risk of death. Even if diagnosed early, there are great chances of recovery and even cure. Treatment is usually long, not less than a year. One of the most critical phases is chemotherapy, which, although very effective, its side effects are generally perverse: frequency of nausea, vomiting and mucositis, somatic changes (alopecia, weight gain, ovarian failure, hormone reduction - testosterone and estrogen causing menopause and vaginal atrophy -, various lows, such as libido, vaginal lubrication, anorgasmia and dyspareunia, affecting sexuality. Inconvenient evaluative judgments emerge for the patient, including the feeling of pity. Even more sinister is the feeling of finitude in the face of the devastating disease [6,9].

In the face of such nefarious problems, it is never too much to insist on prevention, nor is it too much to insist on good eating and behavioural habits, physical exercise, and a regulated life with sleep patterns, which are basic concepts of Public Health. We insist that mammography is essential in this prevention, although its sensitivity is approximately 85% and can be reduced to 50% in very dense breasts. In this sense, any alteration identified in the exam must be described according to the Breast Imaging Reporting in the Data System (BI-RADS), "a system created to standardize the reports of breast diagnostic exams regarding the terms used, report creation and recommendation of conduct, which facilitates

communication between the multidisciplinary team that assists the patient" [9]. Also, images can be improved using the Threshold technique.

2.0 METHODOLOGY

It was necessary to obtain a previously classified image bank to evaluate the proposed methodology, with suspicious regions already segmented. Heath, et al.. (2000) [11] provided updated and standardized version of the Digital Database for Screening Mammography (DDSM). The CBIS-DDSM (DDSM Cured Breast Images Subset) dataset includes uncompressed images, data selection, and curation by trained mammographers. In this work, the tool used to extract suspicious candidate regions in mammography images was the Fiji software, a free software distribution of the *ImageJ* project. *Fiji* is a software focused on the analysis of biological images. It relies on the combination of powerful software libraries with a wide variety of scripting languages that allow rapid prototyping of image processing algorithms.

2.1 Digital Image

Guzman et al, (2013) [12] define an image as a two-dimensional function $f(x, y)$ of light intensity, where x and y denote the spatial coordinates of a point, and the value of the function f (Figure 1) is proportional to the brightness of the image at that point (monochrome image). A digital image is a function $f(x, y)$ discretized in spatial coordinates and brightness. This function produces luminance and reflectance at each point (x, y) .

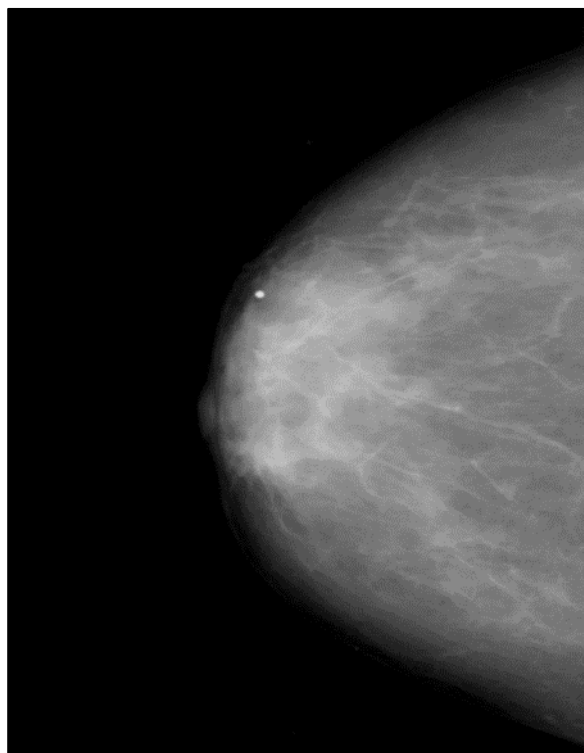


Figure 1: Mammography

2.2 Monochrome Image

A monochrome image contains pixels with only one shade of grey (grayscale). Every scale has a minimum and maximum value. In the case of grayscale (with 8 bits), pixels that approach zero are the darkest pixels, while those that approach the maximum value minus one ($L-1$) are the lightest pixels [13].

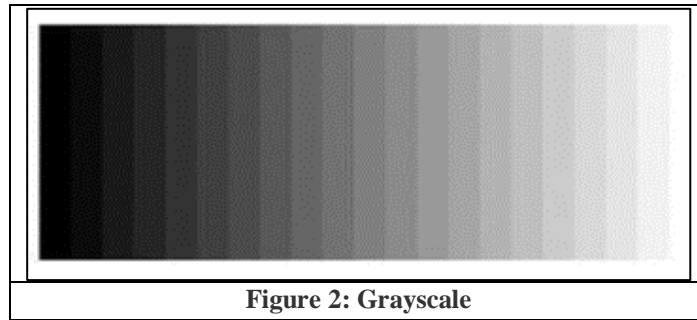


Figure 2: Grayscale

2.3 Thresholding

It is a technique that separates the regions of an image when it has two classes (the background and the object). Because Thresholding produces a binary image as an output, this process is often called binarization. Like the monochromatic image, a binary image is a two-dimensional matrix with only two values. Sometimes they are called logical images: black corresponds to the value 0 and white corresponds to the value 1 [14].



Figure 3: Binary Image

2.4 Neural Networks

Neurophysiologist McCulloch and mathematician Walter Pitts in 1943 modelled simple artificial neural networks using electrical circuits. According to McCulloch and Pitts (1943) [15], a neuron can be represented through binary logic (0 or 1). Artificial neural networks emerged from the search to solve problems analogous to the brain.

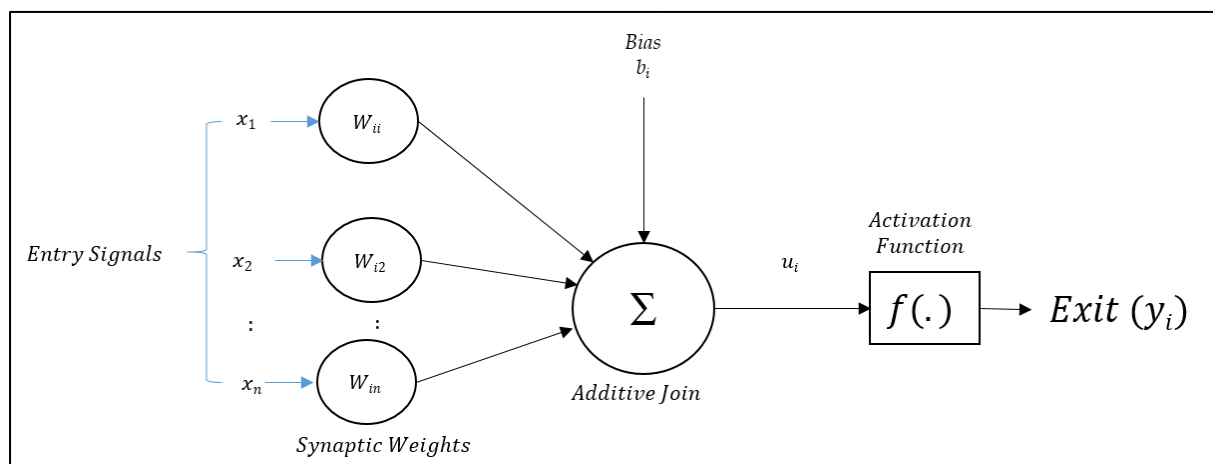


Figure 4: Model of an Artificial Neuron

2.5 Random Forest

Random Forest is a classifier that consists of a collection of decision trees. The idea is that if a tree is good, a forest should be even better, as long as there is enough variety within it. The most exciting thing about a random forest is how it creates randomness from a standard dataset.

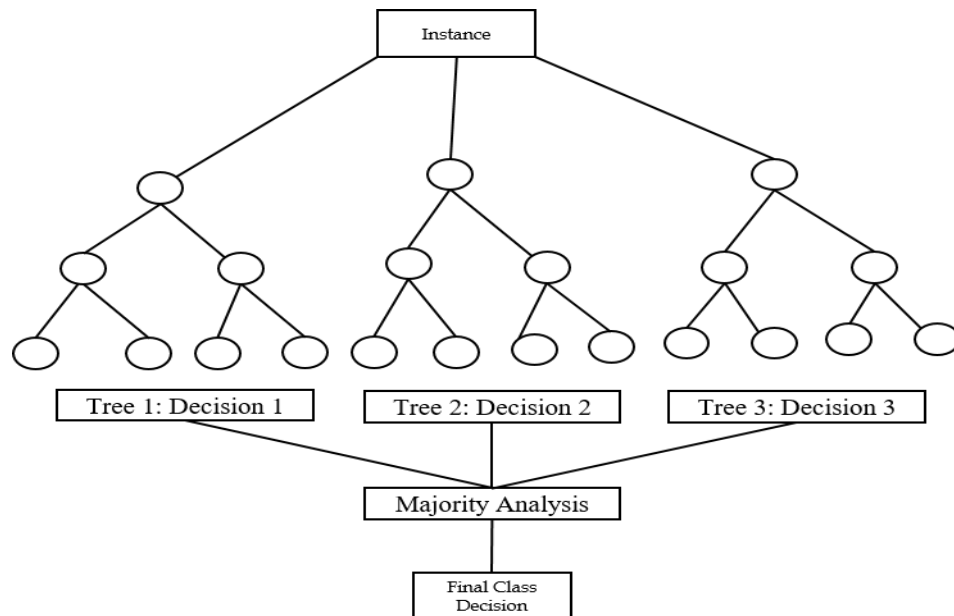


Figure 5: Model of a Random Forest

3.0 RESULTS AND DISCUSSION

The segmenting of the candidate suspect regions is based on a feature vector generated for each pixel in the image. In this way, it is possible to obtain a pattern of attributes that can distinguish calcifications. This method uses various image processing filters on the exam image, generating an image of multiple channels. Each image obtained by applying the filters will be used as an attribute for training a machine learning algorithm [16].

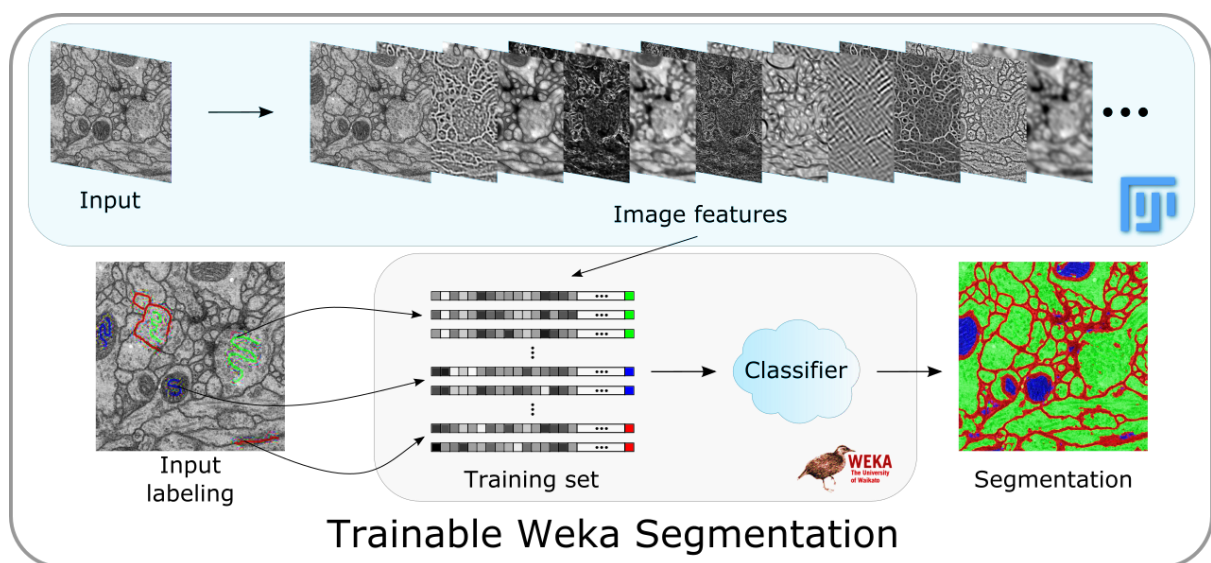


Figure 6: Vector of Attributes of an Image Applied to an Algorithm [16]

As shown in Figure 6, the machine learning algorithms are provided by the Waikato Environment for Knowledge Analysis (WEKA) data mining and machine learning toolkit [17].

The image filters that best meet the extraction of calcification characteristics in mammographic images were selected through this Fiji plugin. Figure 7 demonstrates the selection of filters and the learning algorithm used in this study.

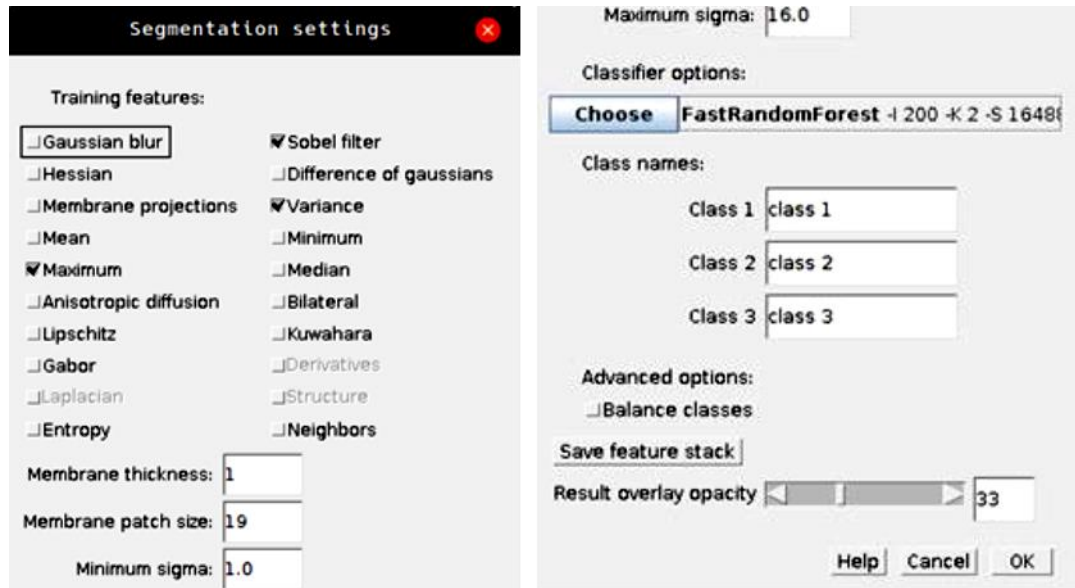


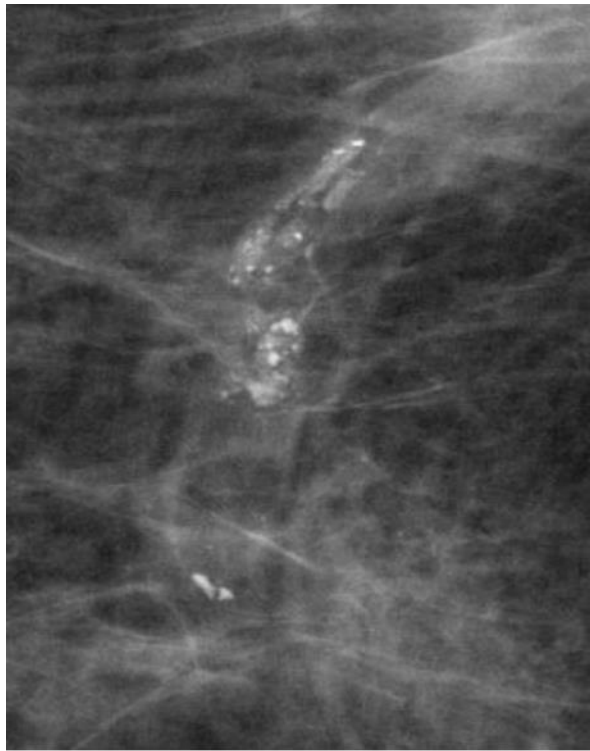
Figure 7: Screen of Parameters by Segmentation with Random Forest

Three filters were selected, an edge detector, variance and maximum value. For each applied filter, new images are generated, making changes to the attributes required for each feature extractor filter. Therefore, the plugin applies different settings for each chosen filter.

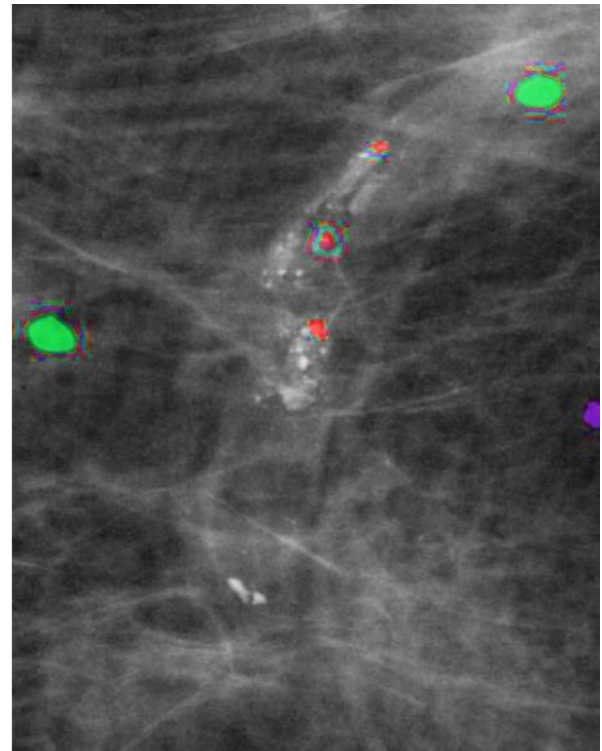
In addition to the filter definition settings and the learning algorithm, three classes were defined for the image segmentation result.

Figure 8b represents the classes in three colours: red is the sample regions of calcification; the green samples represent other regions of the image; and, finally, in purple, the darkest regions to be ignored.

Once the training samples are defined, the feature vectors will be generated; the values of the sample regions will be used as training data for the Random Forest algorithm. After the training phase, the prediction will occur for each pixel of the image that will be segmented.



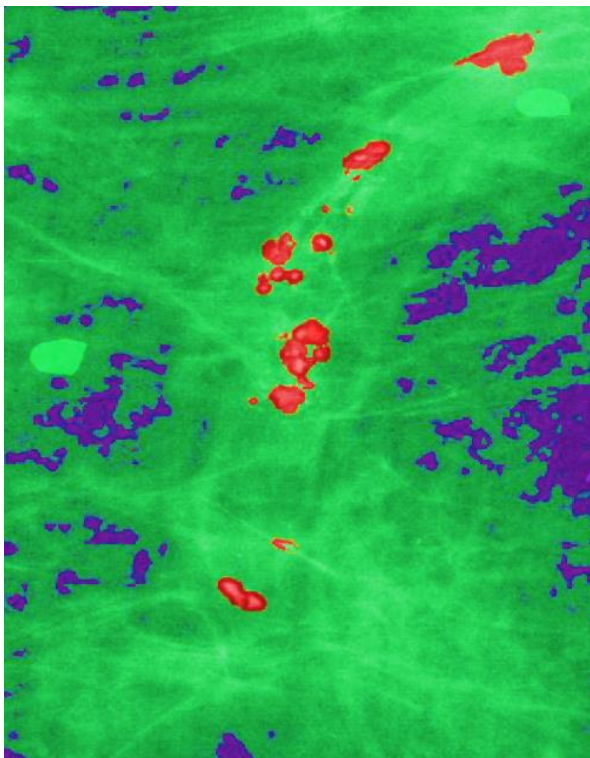
(a)



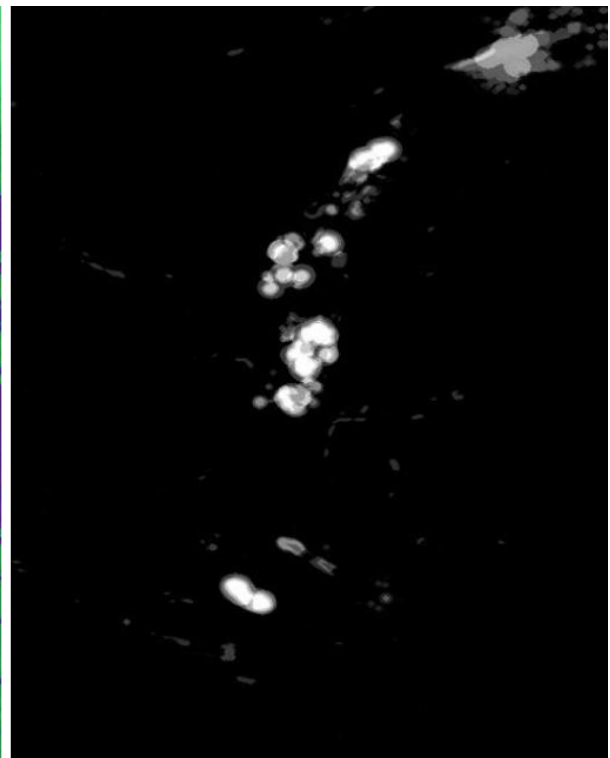
(b)

Figure 8- (a): Original Image;

(b): Image with the Classes



(a)



(b)

Figure 9 (a): Prediction using Random Forest

(b): Random Forest Probability Map.

The images resulting from the segmentation algorithm can be seen in Figure 9: the first represents the three previously defined classes, and the second is the probability map of the candidate suspect regions.

The algorithm applied for image segmentation returns the probability for each image pixel to belong to the first defined class, calcifications; in this way, Figure 9 presents lighter tones for a greater possibility of the suspicious region. The probability of each pixel in the image has a value between 0 (zero) and 1 (one), so 1 (one) corresponds to one hundred per cent of accuracy.

To evaluate using the same image feature extraction method, segmentation tests were performed using neural networks as a learning algorithm. The image filters to create the vector of features for each image pixel were the same applied to the Random Forest algorithm, with only the learning technique being different.

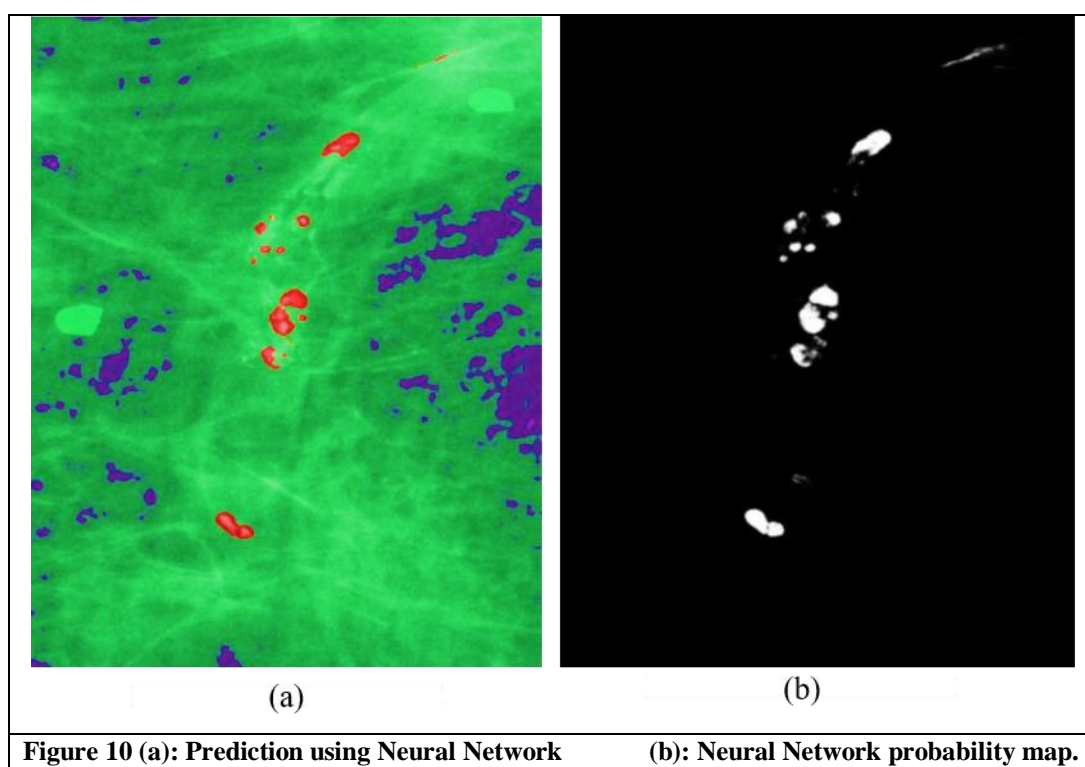


Figure 10 (a): Prediction using Neural Network

(b): Neural Network probability map.

The neural network presented an excellent result in this segmentation methodology. Note that the probability map image presents values closer to white; thus, less noise from suspicious regions. The segmentation results, shown in Figures 23 and 24, demonstrate that it can highlight suspicious regions. Through these probability images, it is possible to detect the contours of the regions; in this way, cuts of these segmented regions can be performed, and, later, an automatic classification of the targets can be performed using a learning algorithm.

4.0 CONCLUSIONS

In this study, the relevance of the intelligent segmentation technique associated with the mammography exam in breast cancer diagnosis was evidenced due to the efficiency of image processing and identification of the suspicious region. In this sense, the research's initial objective was to verify that this technique facilitates the interpretation of findings by the physician, greatly helping the early diagnosis of breast cancer.

The results were satisfactory in the test of the segmentation methodology, which seeks to solve problems in a similar way to the human brain, which, in the contour of suspicious segmented regions, made possible cuts and, from which, classify the targets, automatically, with support in a learning algorithm.

The Random Forest algorithm combines several decision criteria to obtain a more accurate prediction, thus allowing more reliable findings, which facilitates physicians' interpretation since suspicious regions are more easily identified.

More studies in this area are needed to improve the technique. Still, it is already possible to dream of a future in which machines and men work together to favour a population, especially when we talk about cancer, a disease whose mortality is still high and little is known about her. Joining forces for early diagnosis is the biggest challenge.

REFERENCES

1. Siddiqui, Ammar & Amin, Junaaid & Alshammary, Freah & Afroze, Eman & Shaikh, Sameer & Rathore, Hassaan & Khan, Rabia. (2020). *Burden of Cancer in the Arab World*. 10.1007/978-3-319-74365-3_182-1.
2. Salim, Elsayed & Moore, Malcolm & Al-Lawati, Jawad & Al-Sayyad, Jamal & Bazawir, Amin & Bener, Abdulbari & Corbex, Marilys & El Saghir, Nagi & Habib, Omran & Maziak, Wasim & Hamdi Cherif, Mokhtar & Seif-Eldrin, Ibrahim & Sobue, Tomotaka. (2009). *Cancer Epidemiology and Control in the Arab World - Past, Present and Future*. *Asian Pacific journal of cancer prevention : APJCP*. 10. 3-16.
3. Fearon, David & Hughes, Sean & Brearley, S.. (2020). *Experiences of breast cancer in Arab countries. A thematic synthesis*. *Quality of Life Research*. 29. 10.1007/s11136-019-02328-0.
4. Najjar, Hesahm & Easson, Alexandra. (2010). *Age at diagnosis of breast cancer in Arab nations*. *International journal of surgery (London, England)*. 8. 448-52. 10.1016/j.ijso.2010.05.012.
5. El-Zaart, Ali & Alhadithi, Moumena. (2013). *Mammogram Images Thresholding For Breast Cancer Detection Using Different Thresholding Methods*. *Journal of Advanced Research in Breast Cancer*. 02. 10.4236/abcr.2013.23013.
6. Farhadihosseinabadi, Behrouz & Hosseini, Fahimeh & Larki, Pegah & Bagheri, Nader & Abbaszadeh-Goudarzi, Kazem & Sepehr, Koushan & Johari, Behrooz & Abdollahpour-Alitappeh, Meghdad. (2018). *Breast Cancer: Risk Factors, Diagnosis and Management*. *Medical Laboratory Journal*. 12. 10.29252/mlj.12.5.1.
7. Mehrgou, Amir & Akouchekian, Mansoureh. (2016). *The importance of BRCA1 and BRCA2 genes mutations in breast cancer development*. *Medical journal of the Islamic Republic of Iran*. *Med J Islam Repub Iran* 2016, Vol. 30:36
8. Romieu, Isabelle & Ferrari, Pietro & Rinaldi, Sabina & Slimani, Nadia & Jenab, Mazda & Olsen, Anja & Tjønneland, Anne & Overvad, Kim & Boutron-Ruault, Marie-Christine et al., (2012). *Dietary glycemic index and glycemic load and breast cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC)*. *The American journal of clinical nutrition*. 96. 345-55. 10.3945/ajcn.111.026724.
9. Hoff, Solveig & Klepp, Olbjørn & Hofvind, Solveig. (2012). *Asymptomatic Breast Cancer in Non-Participants of the National Screening Programme in Norway: A Confounding Factor in Evaluation?*. *Journal of medical screening*. 19. 177-83. 10.1258/jms.2013.012090.
10. Barba, Diego & León-Sosa, Ariana & Lugo, Paulina & Suquillo, Daniela & Torres, Fernando & Surre, Frederic & Trojman, Lionel & Caicedo, Andres. (2021). *Breast cancer, screening and diagnostic tools: All you need to know*. *Critical Reviews in Oncology/Hematology*. 157. 103174. 10.1016/j.critrevonc.2020.103174.
11. Heath, M & Bowyer, Kevin & Kopans, D & Moore, R & Kegelmeyer, P. (2000). *The Digital Database for Screening*

- Mammography. Proceedings of the Fourth International Workshop on Digital Mammography.* 10.1007/978-94-011-5318-8_75.
12. Guzman-Cabrera, Rafael & Guzmán-Sepúlveda, José Rafael & Torres-Cisneros, Miguel & May-Arrioja, D. & Ruiz-Pinales, José & Ibarra-Manzano, Oscar & Avina-Cervantes, Juan & Gonzalez-Parada, Adrian. (2013). *Digital Image Processing Technique for Breast Cancer Detection.* *International Journal of Thermophysics.* 34. 1519. 10.1007/s10765-012-1328-4.
 13. Fang, Yan & Zhao, Jing & Hu, Lingzhi & Ying, Xiaoping & Pan, Yanfang & Wang, Xiaoping. (2019). *Image Classification toward Breast Cancer using Deeply-Learned Quality features.* *Journal of Visual Communication and Image Representation.* 64. 102609. 10.1016/j.jvcir.2019.102609.
 14. Yedjou, Clement & Tchounwou, Solange & Alo, Richard & Elhag, Rashid & Mochona, Bereket & Latinwo, Lekan. (2021). *Application of Machine Learning Algorithms in Breast Cancer Diagnosis and Classification.* *International journal of science academic research.* 2.
 15. Smalheiser, Neil. (2000). *Walter Pitts. Perspectives in biology and medicine.* 43. 217-26. 10.1353/pbm.2000.0009.
 16. Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., & Sebastian Seung, H. (2017). *Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification.* *Bioinformatics*, 33(15), 2424–2426. doi:10.1093/bioinformatics/btx180
 17. Jin, Wei. (2020). *Research on Machine Learning and Its Algorithms and Development.* *Journal of Physics: Conference Series.* 1544. 012003. 10.1088/1742-6596/1544/1/012003.
 18. Senthil, P. V., V. A. Sirusshti, and T. Sathish. "Artificial Intelligence Based Green Manufacturability quantification of a unit production process." *International Journal of Mechanical and Production Engineering Research and Development* 9.2 (2019): 841-852.
 19. Bryndin, Evgeniy. "Development of sensitivity and active behavior of cognitive robot by means artificial intelligence." *International Journal of Robotics Research and Development* 10.1 (2020): 1-11.
 20. Pande, Pravin P., and R. K. Sambhe. "Artificial Intelligence Machining Predictions by Fuzzy GUI For Inconel 718 with PVD Coated Carbide Cutting Tool." *International Journal of Mechanical and Production Engineering Research and Development (IMPERD)* 7.3 (2017): 313-320.
 21. Nawalagatti, Amitvikram, and R. Kolhe Prakash. "A Comprehensive Review on Artificial Intelligence Based Machine Learning Techniques for Designing Interactive Characters." *International Journal of Mathematics and Computer Applications Research (IJMCAR)* 8.3 (2018): 1-10.

